DOCUMENT RESUME

ED 235 462 CS 007 251

AUTHOR Humphreys, Lloyd G.; Davey, Timothy C.
TITLE Anticipation of Gains in General Information: A Comparison of Verbal Aptitude, Reading Comprehension, and Listening. Technical Report No. 282.
INSTITUTION Bolt, Beranek and Newman, Inc., Cambridge, Mass.; Illinois Univ., Urbana. Center for the Study of Reading.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE Jul 83
CONTRACT 400-81-0030
NOTE 26p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Cognitive Processes; Comparative Testing; Elementary Secondary Education; *Individual Differences; Knowledge Level; *Listening Comprehension; Measures (Individuals); *Reading Comprehension; *Reading Research; *Test Reviews; *Test Theory

ABSTRACT

Four tests were chosen for a study that investigated the hypothesis that individual differences in aural comprehension might anticipate individual differences in general information. A composite of eight subtest scores from the Test of General Information (TGI) was used as a measure of students' general knowledge. The Listening section of the Scholastic Test of Educational Progress (STEP) was included to check on earlier findings; the Verbal Aptitude of the School and College Aptitude Test (SCAT) was selected on the grounds that aptitude should anticipate achievement; and the Reading Comprehension section from STEP was chosen on the hypothesis that skill in reading ought to lead to gains in information. Complete data for each of the tests and the information composite were available for grades 5, 7, 9, and 11 and for constant samples of 748 white girls and 655 white boys. Because samples were much smaller, black boys and girls were omitted from this analysis. Separate matrices of correlations were computed for each sex. From each master matrix, three additional matrices were formed of each test and the composite. Results supported the hypothesis. Reading comprehension was most highly correlated with individual differences in general information when all were measured on the same occasion. In contrast, individual differences in estimated true scores of listening were most highly correlated with general information when there were at least 2 years between the occasions of measurement. (HOD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235462

CENTER FOR THE STUDY OF READING

Technical Report No. 282

ANTICIPATION OF GAINS IN GENERAL INFORMATION: A COMPARISON OF VERBAL APTITUDE, READING COMPREHENSION, AND LISTENING

Lloyd G. Humphreys and Timothy C. Davey

University of Illinois at Urbana-Champaign

July 1983

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02238

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

The research reported herein was supported in part by the National Institute of Education under Contract No. NIE-400-81-0030.

CS 007251

EDITORIAL BOARD

William Nagy
Editor

Harry Blanchard
Wayne Blizzard
Nancy Bryant
Pat Chrosniak
Avon Crismore
Linda Fielding
Dan Foertsch
Meg Gallagher
Beth Gudbrandsen
Anne Hay
Patricia Herman
Asghar Iran-Nejad
Margaret O. Laff
Brian Nash
Theresa Rogers
Terry Turner
Paul Wilson

Anticipation of Gains in General Information: A Comparison of Verbal Aptitude, Reading Comprehension, and Listening

Between 1961 and 1967 the Educational Testing Service obtained data for a longitudinal study of academic growth (Hilton, Beaton, & Bower, 1971). Students were tested in a wide sampling of American schools in the 5th, 7th, 9th, and 11th grades. A substantial number of the examinees furnished a full set of scores in all grades. The cognitive measures used were the two subtests from the School and College Aptitude Test (SCAT), the six subtests of the Scholastic Test of Educational Progress (STEP), and the eight scores of the Test of General Information (TGI). Such data are well designed for the application of the cross-lagged correlation (CLC) analysis, popularized by Campbell and Stanley (1963). A data tape was made available to the present senior author for purposes of CLC analyses by Thomas Hilton and the Educational Testing Service. The present report is a continuation of earlier ones written in collaboration with various students. The earlier results that bear on the present problem are briefly summarized.

Atkin, Bray, Davison, Herzberger, Humphreys, and Selzer (1977) conducted a largely traditional CLC analysis of relationships involving each of the 16 test scores mentioned above and a composite formed from the same tests. In samples of white females, white males, black females, and black males they found that Listening, a measure of aural comprehension, clearly differed from the rest. Listening predicted a composite formed from the remaining 15 tests more accurately than the composite predicted Listening over two, four, and six-year time periods. The composite can be considered the

equivalent of a test of general intelligence. Listening was seemingly tapping the causes of intellectual growth more directly than any of the other measures.

We avoided then, and have avoided since, reaching a conclusion from uncontrolled correlational analysis that aural comprehension causes intellectual growth. A neutral statement concerning causation that we have used consistently is that individual differences in aural comprehension anticipate later individual differences in a cognitive composite.

The next study in the series checked on the outcome of the CLC analysis by means of multiple factor analysis. We hoped that we might develop an hypothesis concerning the mechanism underlying the cross-lagged differences from the factor structure. Fitting the intercorrelations of 16 measures, each obtained on four occasions, with a multiple factor model that reflects changes in the true scores of individuals on the factors during development is complex. Humphreys and Parsons (1978), however, obtained a satisfactory fit for 16 measures for the two extreme occasions, grades 5 and 11. The results were congruent with the earlier conclusion that individual differences in Listening anticipated later individual differences on the composite, but the analysis did not suggest any reasons for the finding. Listening at grade 5 was located within the highly oblique space defined by the two vectors representing the general factor in the 16 test scores at 5 and 11. Listening at grade 11 was outside that space and, of course, closer to the general factor at grade 11. Less technically, Listening at grade 5 was equally close to the two general factors whereas Listening at grade 11 was close only to the general factor in the grade 11 measures. Otherwise,

there was no appreciable change in the factor structure of Listening from one grade to the other.

Humphreys and Parsons (1979) subsequently presented a different model that used the properties of the simplex correlation matrix for both the Listening test and the cognitive composite. Their simplex process model provided an estimate of the lag in time between individual differences in Listening and in the composite of from two to four years. They also demonstrated that traditional CLC analysis was vulnerable to differences in the stabilities of the correlations between true scores from one occasion to another. If individual differences on one variable are changing more rapidly than individual differences on a second, a cross-lagged difference in the correlations can be produced with zero lag. Their model controlled for this artifact, although in this instance the true score stabilities for Listening and the cognitive composite were approximately equal. Thus the more traditional CLC analysis for these variables was validated. Humphreys, Park, and Parsons (1979) immediately applied the same model to all of the 16 tests in each of the four sex and race groups used originally. Again the Listening test stood out from the rest in anticipating later individual differences in the printed test composite.

For a complete description of the model the reader is referred to the references cited. In brief, the model starts with diagonal factoring of the intercorrelations of each measure separately over occasions. A lag hypothesis is tested by matching the first diagonal factor, defined by the first occasion for the anticipating variable, with a succession of first diagonal factors extracted from the first to the n'th occasion for the

lagging variable. The form of the cross-correlations between the two variables is determined by the characteristics of the diagonal factors of a simplex matrix of correlations and the lag required for the best match.

Following these two articles that used the simplex process model a critique of the CLC methodology was published by Rogosa (1980). He discussed at length the problem of differential stabilities.[1] He also omitted any discussion of the simplex process model and concluded that the CLC methodology should be completely abandoned. Because the simplex process model was a development from CLC, a reader of Rogosa's critique might infer that the model was defective. It does, however, solve the problem of differential stabilities of the two measures being compared.

Even though the model is basically sound, it has been suggested that the substantive findings are essentially trivial. Carroll and Maxwell (1979) argued that there are extraneous sources of variance, such as differences in reading skills, that affect the measurement of intelligence by means of printed tests. These extraneous sources become progressively less important between grades 5 and 11 for tests requiring reading, but have no impact on Listening. Although blacks had a one standard deviation mean deficit in both Reading Comprehension and Listening, their lags of the composite behind Listening did not differ from those for whites. This suggests that differences in reading did not produce the findings. Nevertheless, we decided that it would be desirable to exercise greater control of reading than had been done earlier and see if similar results would still be obtained. This required selection of a criterion measure that would minimize the contribution to variance of individual differences of reading skill.

Procedure

The TGI test differs from the usual academic achievement test such as those in STEP in that the questions are simple and require merely a factual answer. The reading comprehension load on an examinee for each of the three forms administered in the research (Hilton, Benton, & Bower, 1971) is minimal. Difficulty level is determined by the difficulty of the fact, not by the sentence that asks the question. We decided, therefore, to form a new composite of the eight scores of the TGI, giving unit weight to each. The contribution to the variance of this composite from differential interests is minimized by the heterogeneity of the eight scores: Industrial Arts, Home Arts, Mathematics and Physical Science, Biological Science, Music and Art, History and Literature, Recreation and Entertainment, and Government and Public Affairs.

We chose three other tests from the original set of 16 for the analysis of the hypothesis that individual differences on each might anticipate individual differences in general information. Listening of STEP was included to check on earlier findings and on the interpretation of Carroll and Maxwell. Relative retardation in reading in connection with the use of printed tests represents an uninteresting causal explanation, but is not plausible if the reading required is well below grade level. Verbal Aptitude of SCAT was selected on the grounds that aptitude should anticipate achievement. Because the criterion was a verbal information test, Quantitative Aptitude was not indicated. Reading Comprehension from STEP was chosen on the reasonable hypothesis that skill in reading ought to lead to gains in information.

Complete data for each of the three tests and the information composite were available in grades 5, 7, 9, and 11 for constant samples of 748 white girls and for 655 white boys. Because samples were much smaller, black boys and girls were omitted from this analysis. Separate 16 x 16 matrices of correlations were computed for each sex. From each master R matrix we formed three 8 x 8 matrices of each test and the composite. In each of the latter matrices there are intercorrelations over four occasions for each variable and 16 cross-correlations, four synchronous and 12 cross-lagged, between the two variables.

The next step was to fit the simplex process model to each of the 8 x 8 matrices. In doing this we departed from the procedures used originally in obtaining stabilities of true scores and in estimating common true score variance of the two measures being compared. With respect to stabilities one can obtain a fit to the simplex for four occasions by setting the three correlations for the adjacent occasions equal to each other, setting the four reliabilities equal, or allowing both to vary somewhat.[2] Earlier we opted for equal stabilities. This time we tried all three. With respect to estimating common true score variance we abandoned the direct estimation of obtained score communalities and conversion to a proportion of true score variance because we discovered a conceptual flaw in our procedure. We substituted a regression estimate that utilized only one degree of freedom from the 4 x 4 matrix of cross-correlations between a given test and the information composite. Once the expected values of the cross-correlations based on perfect communality were found, we used the regression of the observed correlations at all grades on the former values to obtain the expected cross-correlations when communality is less than unity.

Our correction for specificity assumes that communalities do not change from one grade to another. This may not be accurate, but it does minimize the amount of information taken from observations. More accurate fits were obtained by estimating communalities for each grade separately, and without using all available degrees of freedom, but the procedure we adopted is more conservative.

Results

Table 1 presents the two master R matrices. Correlations for girls appear above the diagonal, for boys below. The tests are identified by the first letter of the title. C is used for the unit-weighted information composite. Subscripts indicate the four different grade levels. Each table of correlations is based on a constant sample so that high school dropouts are not included. Nevertheless a relatively wide range of talent is represented as a result of the sampling in the growth study as described by Hilton et al. As a result the contribution to variance of the general factor is high. A high level of stability from grade 5 to 11 is also apparent, but looking only at cross-lagged differences among those correlations based on observed scores one finds that Listening predicts the composite more accurately than the composite predicts Listening.

Insert Table 1 about here.

The three different assumptions concerning stabilities and reliabilities actually converged on very similar outcomes for the cross-correlations between the individual tests and the composite. Stated in terms of lag of the information composite behind the individual tests, the lag appeared

to be two to four years behind Listening, zero to two years behind Reading Comprehension, and zero with respect to Verbal Aptitude. Two of the three assumptions provided equally good fits of a simplex matrix to the intercorrelations of the variables individually over the four occasions. Equal reliabilities provided an additional degree of freedom for the chi square test, but provided significantly poorer fits in six of eight comparisons and wide variation in stabilities. Equal stabilities provided better fits, but produced wide variation in reliabilities. The assumption that allowed both reliabilities and stabilities to vary produced equivalent fits and plausible stabilities and reliabilities. To conserve space only data from this last, the more plausible assumption, will be presented.

Table 2 contains the chi squares for the goodness of fit to the theoretical simplex matrix of the observed intercorrelations over occasions for each of the four variables. Also included are the estimates of reliabilities and stabilities. The reliabilities of the first occasion of measurement was tied to the reliability of the second and the fourth to the third. Data are presented separately for boys and girls.

Insert Table 2 about here.

With one degree of freedom three chi squares are significant with p less than .05 and two with p less than .01. Nevertheless, the largest discrepancies between observed and expected are only slighlty larger than .02 in the poorer fits. The sample sizes mitigate the impact of the significant chi squares. Reliabilities are about as expected. Verbal Aptitude and the information composite are somewhat more reliable than

Listening and Reading. The stabilities of the composite between grades 9 and 11 are inflated by the fact that the same form of the TGI was used in both grades. Two different forms were used in grades 5 and 7. Stabilities seem to level off after grade 7 for Verbal Aptitude and Reading, but there is an inexplicable drop from 9 to 11 for Listening. It must be remembered, however, that the assumption of equal stabilities from one occasion to another provides an equally good fit and, necessarily, different reliabilities as well.

In fitting the cross-correlations between tests and the composite we examined the four possible lags made possible by the four occasions of measurement: namely, zero, two, four, and six years. It would have been possible to add lags in the other direction from information to the tests, but there is no basis in theory or in the data to do so. For each hypothetical lag we computed two separate sums of squared deviations as descriptive statistics of goodness of fit, one with the synchronous correlations included and one with them excluded. Humphreys and Parsons had argued that there was a reasonable possibility of correlated measurement error among synchronous correlations. These descriptive statistics are presented in Table 3.

Insert Table 3 about here.

It is seen that a minimum sum of squares is reached at zero lag for Verbal Aptitude and Reading and at two years for Listening. These minima are independent of whether the synchronous correlations are included or excluded, but the mean of 16 squared deviations tends to be somewhat

larger than the mean of 12. It is also evident that the squared deviations at the point in time of best fit are close to being the same size for the three tests and for the two sexes.

Based on the size of the indices of goodness of fit at lags adjoining the minima, the lag of zero for Verbal Aptitude and the composite appears to be highly accurate. In contrast, the lag for Reading might be nonzero and the lag for Listening might be greater than two years. At one stage we estimated communalities separately at each grade level of the test. When this is done, a lag of four years provides a better fit for listening, but the separate communalities had no effect on the estimate of the length of the lag for either Verbal Aptitude or Reading. We emphasize that the zero lag for the former is not in line with expectations based on notions about aptitude, but neither is the very high communality between verbal comprehension and general information congruent with those same notions. The near functional equivalence belies the difference in the names.

Table 3 includes all of the data required to support a conclusion that Listening differs from Verbal Aptitude and Reading Comprehension in anticipating individual differences in general information by at least two years. However, the reader might be interested in an example of the computations intervening between estimation of reliabilities and stabilities and the descriptive indices of goodness of fit of the model to the observed cross-lagged correlations. The successive matrices for the girls for the Listening test and a two-year lag will make the summaries of the results more meaningful.

Table 4 contains the diagonal factors required for the model fitting. Factor 1 for Listening is defined by the true scores at grade 5, but

Factor I for the information composite is defined by the true scores at grade 7. This represents the lag of two years. The hypothesis that these and the subsequent factors, pair by pair, are identical except for a communality difference is tested by a matrix multiplication. This step disregards the factor in Listening at grade 11 and the factor in the composite at grade 5. These latter factors are required only for zero lag.

Insert Table 4 about here.

Table 5 provides on the left the product of the Listening factor matrix and the transpose of the composite factor matrix. In the center is the same matrix after allowance is made for the less than perfect communality between the matched factors. On the right is the observed matrix of cross-correlation. For this example, the regression coefficient (.869) used to produce the center matrix from the one on the left was based on the 16 observed correlations. That is, the synchronous correlations that are possibly inflated by correlated measurement error are included. That the correlation at grade 5 does not appear inflated can be interpreted as a rejection of the assumption that communalities are stable at all grades.

Insert Table 5 about here.

Not only does the constant communality overestimate the synchronous correlation at grade 5, but it also produces the single largest residual. Overall, however, the fit is quite good. It depends primarily on information from the intercorrelations of the separate variables over occasions.

Only one item of information was obtained from the 16 correlations we were attempting to fit.

Discussion

No matter how we analyze these data we reach the same conclusion: namely, individual differences in Listening, a measure of aural comprehension of the English language, anticipate later individual differences in general information. Earlier research used a different composite consisting of additional, more complex components than in the present research and found that the lag between individual differences in Listening and the cognitive composite was from two to four years. We now conclude that the lag between individual differences in aural comprehension and in amount of general information is at least two years. We are convinced that the phenomenon is real.

It also seems reasonable to believe that the causes are not superficial ones. The questions on the TGI have to be read, but the reading task is a relatively simple one at each grade level. Students who are relatively high in aural comprehension of language in the earlier grades are the ones who are more likely to increase their fund of general information in the following two or more years than those who are proficient in reading or high in so-called verbal aptitude. More fundamental causes should be considered than mere facility in reading printed test material.

To conclude that the causes are not superficial seems sound, but to go beyond that is more difficult. An early hypothesis was that the Listening test tapped attentive behaviors more directly than the printed tests and that these behaviors facilitate the acquisition of academic

skills and information. This still seems reasonable. The format of the Listening test in STEP follows that of the Reading Comprehension test, but the paragraphs and questions are presented orally. Examinees are not able to go back and forth from question to paragraph. They must comprehend the thought expressed and relate it to the subsequent question based on only a single presentation of each. Attention cannot be allowed to fluctuate. These test taking behaviors presumably generalize to classroom behavior. On their face they are important constituents of an effective teaching-learning situation.

In interpreting these data the reader must clearly distinguish between anticipation and prediction. All three tests predict scores on the information composite two, four, and six years later. A multiple regression composite of the three tests would have a higher correlation with the information criterion than any one of the three. Listening may add very little to the prediction of the information criterion when the data are observed scores and little more when the correlations are corrected for attentuation.

Verbal Aptitude especially is a more effective predictor than Listening. Individual differences in estimated true scores of Verbal Aptitude and Reading Comprehension, however, are most highly correlated with individual differences in general information when all are measured on the same occasion. In contrast individual differences in estimated true scores of Listening are most highly correlated with general information when there is a lag of at least two years between the occasions of measurement. True scores of individual examinees on all of these measures change in relative

position in the distribution from one occasion to another. Change is greater the longer the interval between occasions. We can only guess at present concerning the causes for these changes, but the Listening test is reflecting the causes of change two or more years earlier than Verbal Aptitude or Reading Comprehension.

There is only one way in which the possible importance of attentive behaviors can be tested with high confidence, but an experimental attack on the problem is difficult in several ways. One must be able to modify experimentally attentive behaviors in classrooms, to maintain these behaviors over a two-year period, to assign classrooms at random to experimental and control conditions, and to obtain sufficiently large numbers of classrooms to obtain dependable results.

In the light of present dependable methods of increasing the ability of students in our schools to read, write, and acquire information, one might be willing to take a chance on less dependable information. Anticipation falls short of a cause and effect sequence, but the development and experimental use of curricula, starting at least as early as the kindergarten year, stressing aural comprehension seems a reasonable gamble. Teachers can also be trained to develop attentive behaviors at all grade levels through the use of selective reinforcement. Durkin's study (1982) of superior black readers is highly pertinent to this discussion. She found that the one common element in the backgrounds of successful black readers, defined as reading at or beyond grade level, was their exposure to a substantial amount of oral reading during the preschool period. As with all research conducted without control of variables and random

assignment to treatments, other causal explanations are possible, but children who had mothers, grandmothers, aunts, or older siblings who read to them were the successful readers.

Footnotes

[1]Rogosa errs, as do many others, when discussing structural equations. He omits the important difference between true and observed scores. Equations are presented with symbols indicating population parameters, and a careless reader assumes that sample statistics as estimates of the population parameters can be inserted in the equations. A theoretical underlying process or latent trait does not change, however, as the measures that assess it vary in reliability. It is essential to have estimates of true scores. The approach through factor analysis (LISREL) avoids the dependence on observed correlations in inferences about either causation or anticipation. Because the simplex process model requires reliabilities and true score stabilities, it is also a latent trait model. This is contrary to Bentler's (1980) classification of it as a manifest variable model. True score is the most elementary latent trait. It can be estimated with less ambiguity than a factor. The latter is always, to some degree, dependent on the selection of the three or more tests required.

[2]In a four-variable simplex only the reliabilities of occasions two and three and thus the true score stability between two and three are determinate. An assumption is required before the remaining parameters can be estimated.

References

Atkin, R., Bray, R., Davison, M., Herzberger, S., Humphreys, L., & Selzer, U. Cross-lagged panel analysis of sixteen cognitive measures at four grade levels. Journal for Research in Child Development, 1977, 48, 944-952.

Bentler, P. M. Multivariate analysis with latent variables: Causal modeling. Annual Review of Psychology, 1980, 31, 419-456.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand McNally & Co., 1966.

Carroll, J., & Maxwell, S. Individual differences in cognitive abilities. Annual Review of Psychology, 1979, 30, 603-640.

Durkin, D. A study of poor black children who are successful readers (Reading Education Report No. 33). Urbana: University of Illinois, Center for the Study of Reading, April 1982.

Hilton, T., Beaton, A., & Bower, C. Stability and instability in academic growth--a compilation of longitudinal data. Final Report, USOE, Research No. 0-0140, August 1971.

Humphreys, L., & Parsons, C. Search for a common factor model to describe a cross-lagged correlation difference. Applied Psychological Measurement, 1978, 2, 257-267.

Humphreys, L., & Parsons, C. A simplex process model for describing differences between cross-lagged correlations. Psychological Bulletin, 1979, 86, 325-334.

Humphreys, L., Park, R., & Parsons, C. Application of a simplex process model to six years of cognitive development in four demographic groups. Applied Psychological Measurement, 1979, 3, 51-64.

Rogosa, D. A critique of cross-lagged correlation. Psycholgoical Bulletin, 1980, 88, 245-258.

Table 1

Intercorrelations of Four Variables Each

Measured on Four Occasions

	R_5	L_5	V_5	C_5	R_7	L_7	V_7	C_7	R_9	L_9	V_9	C_9	R_{11}	L_{11}	V_{11}	C_{11}
R_5		71	76	64	75	67	76	62	66	56	72	60	69	57	70	60
L_5	68		73	66	71	74	76	67	65	69	74	69	67	63	73	70
V_5	80	65		79	73	71	85	75	64	60	79	72	68	58	76	72
C_5	72	61	80		64	62	77	78	55	54	72	72	59	53	69	72
R_7	72	62	68	63		74	80	67	75	67	79	69	72	62	77	69
L_7	67	75	66	62	69		78	69	68	75	79	74	69	70	77	74
V_7	78	68	84	78	76	75		78	71	70	88	78	74	64	85	78
C_7	68	64	73	83	66	68	77		60	63	77	79	63	59	74	78
R_9	69	62	66	62	75	68	72	64		70	75	65	72	66	74	67
L_9	55	67	58	53	61	74	64	57	69		74	72	65	70	74	70
V_9	72	68	76	73	71	75	85	75	76	69		80	75	67	87	81
C_9	68	66	73	78	65	73	79	82	71	66	82		65	63	78	85
R_{11}	67	62	65	61	68	66	70	63	75	65	72	70		67	80	72
L_{11}	56	62	53	50	51	64	58	53	60	65	60	58	67		74	67
V_{11}	71	68	75	71	71	73	81	73	73	71	83	80	77	70		82
C_{11}	66	66	71	74	64	72	77	79	69	67	79	88	73	64	83	

Correlations for 748 girls are above, for 655 boys below, the diagonal; correlations are rounded to two places and decimal points are omitted.

Table 2

Chi Squares of Goodness of Fit of the Simplex Matrix, Reliabilities[a], and Stabilities[b] of Each of the Four Measures

	Composite		Reading Comprehension		Verbal Aptitude		Listening	
	Male	Female	Male	Female	Male	Female	Male	Female
x^2	.006	.289	7.732**	12.375**	3.033	.421	4.376*	.048
$r_{(5)(5)}$	.879	.853	.766	.815	.925	.943	.813	.819
$r_{(7)(7)}$	.879	.853	.766	.815	.925	.943	.813	.819
$r_{(9)(9)}$	.919	.857	.809	.731	.858	.894	.731	.764
$r_{(11)(11)}$	.919	.857	.809	.731	.858	.894	.731	.764
$r_{(5)(7)}$	.947	.919	.937	.914	.908	.900	.919	.909
$r_{(7)(9)}$	.918	.925	.947	.962	.952	.958	.954	.953
$r_{(9)(11)}$	.952	.987	.931	.985	.965	.970	.885	.922

[a]Estimated correlations between parallel measures at a given grade level, e.g., $r_{(5)(5)}$

[b]Estimated correlation betwen true scores for adjacent grade levels, e.g., $r_{(5)(7)}$

*$p < .05$

**$p < .01$

Table 3

Sums of Squared Deviations Between Observed and Expected Cross-Correlations for Each Test and the Information Composite

Lag in Years	Reading Comprehension		Verbal Aptitude		Listening	
	Males	Females	Males	Females	Males	Females
			Based on 16 Correlations			
0	.01421	.01481	.01402	.00798	.04180	.04009
2	.01882	.02131	.04609	.04265	.01261	.01103
4	.04468	.06322	.12068	.10891	.02388	.02902
6	.05231	.07268	.14917	.12068	.03762	.03561
			Based on 12 Correlations (Synchronous Excluded)			
0	.01153	.00916	.00834	.00178	.03121	.03303
2	.01522	.01798	.04369	.04056	.00740	.00755
4	.03117	.05281	.10693	.09808	.01489	.01940
6	.03976	.06208	.13195	.10716	.02825	.02465

Table 4

Diagonal Factors in Listening and in the

Composite for the Hypothetical Two-Year Lag

N = 748 Girls

	Factors in Listening Occasions			Factors in Composite Occasions		
	5	7	9	7	9	11
Grade 5	863	000	000	849	-011	019
Grade 7	863	274	000	923	000	000
Grade 9	795	245	268	858	346	000
Grade 11	730	241	243	844	349	371

Table 5

Expected and Observed Cross-Correlations Between

Listening and the Information Composite

$\underline{N}$ = 748 Girls

	Expected, h^2 = 1.00				Expected, h^2 = .869				Observed			
	C_5	C_7	C_9	C_{11}	C_5	C_7	C_9	C_{11}	C_5	C_7	C_9	C_{11}
L_5	77	84	78	76	68	74	69	68	66	67	69	70
L_7	69	76	84	83	62	67	74	73	62	69	74	74
L_9	65	70	77	80	54	62	68	71	54	63	72	70
L_{11}	60	64	71	74	53	57	63	65	53	59	63	67

Correlations are rounded to two places and decimal points are omitted.